

Universal relations in the self-assembly of proteins and RNA

D. Thirumalai¹

¹*Institute for Physical Science and Technology,
University of Maryland, College Park, Maryland 20742, USA*

Concepts rooted in physics are becoming increasingly important in biology as we transition to an era in which quantitative descriptions of all processes from molecular to cellular level are needed. In this essay I discuss two unexpected findings of universal behavior, uncommon in biology, in the self-assembly of proteins and RNA. These findings, which are surprising, reveal that physics ideas applied to biological problems ranging from folding to gene expression to cellular movement and communication between cells might lead to discovery of universal principles operating in adoptable living systems.

The fascination of physicists and mathematicians with biology can be traced back to at least the turn of the twentieth century. Two remarkable books, one on morphogenesis grappling with patterns on animals and plants by Thompson [1] and the other by Schrodinger [2] outlining how the principles of statistical mechanics and stochastic fluctuations govern life itself, illustrate beautifully the need for a quantitative perspective on biology. These works, which have inspired many scientists to delve into the complexities of biology, continue to be relevant to this day. The need for molecular explanations of a number of phenomena in biology also began early in the 1900s when Christian Bohr, the father of Niels Bohr, discovered the impact of pH (Bohr effect) on oxygen binding to hemoglobin, a problem that has dominated research in molecular biophysics. Although the use of physics ideas in biology is not new the intensity with which physical principles are being applied to living systems is unprecedented, drawing researchers from a variety of backgrounds. As a result, there is hardly any subfield of biology in which the presence of physics ideas is not felt.

My interest in using the tools of physics to understand biology began in the late 1980s when I got interested in protein folding [3]. Dana Honeycutt and I had worked on the effects of randomly placed obstacles on the shapes of homopolymers, a prelude to my interests in crowding effects on proteins and RNA that came years later. After we finished that work, Dana asked me if I had any suggestions on what we could do next, and he proposed that we tackle the dynamics of a polymer molecule in a random environment. I talked him out of this because the problem was (and is) difficult. I also felt that the results might be of interest to only a small number of scientists. Instead, I suggested that we look into protein folding. It is generally assumed, thanks to the groundbreaking experiments by Anfinsen [4], that the number of folded states of a protein is indeed small (in fact unique) implying that it is likely that the functionally competent state corresponds to the lowest (or near lowest) free energy state

amidst all the exponentially large number of conformations a polypeptide chain can adopt. I wanted to understand why proteins have to reach the lowest free energy minimum (Anfinsen hypothesis) in order to execute their functions. I theorized to Dana that proteins could have countable number of low free energy minima in which they could adopt similar structures without compromising their ability to carry out their functions. In other words, a functioning protein could be metastable. Using a minimal model of a protein, one that has been adapted and modified by many researchers, Honeycutt and I [3] argued that the functional state of a protein could be metastable, a notion which was not inconsistent with experiments but nevertheless was considered heretical. Upon publication of our paper [3], Kaufmann wrote us that he too had similar ideas and discovered that biologists were not too receptive to them. He did alert us to a series of experimental papers reporting that functional states (plural!) could be metastable. Interestingly, more recently it has been demonstrated that the folded state of mammalian prions may well be metastable [5] with the more stable conformation being aggregation prone, and hence deleterious. The concept that folded states could be metastable seems more easily accepted for ribozymes [6].

Prior to the announcement of the metastability hypothesis in [3], papers demonstrating uses of statistical mechanics of disordered systems [7] and polymer theory [8] as a way to describe the self-assembly of proteins appeared. The modern perspectives have literally transformed the field of protein folding research, a trend that continues unabated to this day. Thus, time was ripe to produce a fresh perspective on the self-assembly of proteins and RNA. We were the first to use physics concepts to describe the complex pathways in the folding of RNA [9]. Our initial work got me hooked and my research group has since then focussed on using physics concepts to describe a number of problems in biology. Here, I will give two example where we unexpectedly discovered universal relations

in protein and RNA folding using ideas rooted in polymer physics and glasses [10], which at least to me is a surprise given the tremendous emphasis on specificity placed by biologists.

Length dependence on folding cooperativity and collapse transition: Single domain proteins undergo a remarkably cooperative transition from the unfolded to a folded state when the conditions for folding become favorable, for example by lowering the temperature. The folding reaction, occurring at T_F may be viewed as a phase transition. Single domain proteins are finite-sized, with the number of amino acid residues in the majority of the experimentally characterized proteins containing less than about 100 residues. Thus, we expect finite-size effects to play an important role in the folding phase transition. Because proteins are polymers, with water being a poor solvent for these systems, we expect that in the process of folding proteins should also undergo a collapse transition, at a characteristic temperature, T_θ . Based on theoretical arguments and precise numerical results for protein-like models we established sometime ago that efficient folding occurs if $T_\theta \approx T_F$ [11], a prediction that has only recently been fully validated [12]. The extent of cooperativity in this transition can be assessed using a dimensionless quantity, $\Omega_c = \frac{T_F}{\Delta T} \left| \frac{df_{NBA}}{dT} \right|$ where f_{NBA} is the fraction of molecules in the NBA at T , ΔT is the full width at half maximum of $\left| \frac{df_{NBA}}{dT} \right|$. In the expression for Ω_c , $\left| \frac{df_{NBA}}{dT} \right|$ is evaluated at T_F . We showed that $\Omega_c \sim N^\zeta$ [13] where N is the number of amino acid residues in a protein and $\zeta = 1 + \gamma$ with γ being the exponent that characterizes the divergence of the susceptibility at the critical point for a n -component ferromagnet with $n = 0$, corresponding to a self-avoiding walk - a reasonable model that describes the global properties of unfolded proteins. An accurate numerical estimate based on a fifth order ϵ expansion of ϕ^4 field theory for a polymer gives $\gamma \approx 1.22$, leading to the conclusion that $\Omega_c \sim N^{1.22}$. We showed [13] that experiments on a number of proteins are in accord with this *universal* prediction.

The rationale for expecting the universal behavior for Ω_c goes as follows: (1) By analogy with magnetic systems we can identify f_{NBA} as an order parameter that distinguishes between the folded and unfolded states, and hence $T \frac{df_{NBA}}{dT}$ can be associated with "susceptibility" with T being the ordering field. (2) Camacho and I [14] showed that the collapse transition at T_θ for finite N could be second order while the folding transition at T_F is (weakly) first order. The condition $T_\theta \approx T_F$ implies that the folding transition itself could have tricritical character, thus tidily explaining the marginal stability of proteins. Therefore, the critical exponents that control the behavior of the polypeptide

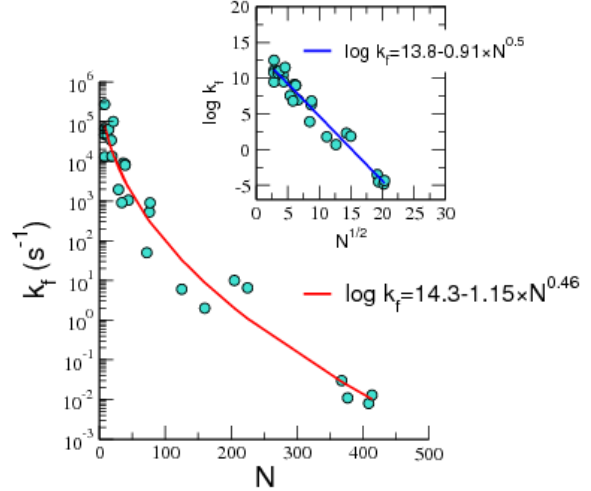


FIG. 1: Dependence of the folding rate of RNA on N . The solid line is the theoretical fit with $\beta \approx 0.46$ as a free parameter. The inset on top shows that the quality of the fit, over seven orders of magnitude, is excellent with β fixed at the theoretical value of 0.50. The figure is adapted from [16].

chain at T_θ should manifest themselves in the folding phase transition. Given that susceptibility scales as N^γ and $\frac{\Delta T}{T_F} \sim N^{-1}$ [13], it follows that $\Omega_c \approx N^\zeta$.

Folding rates of proteins and RNA scale as $e^{\sqrt{N}}$: Using theoretical arguments, whose genesis is in the dynamics of activated transitions in supercooled liquids [10], I suggested that folding rates of proteins can be written as, $k_F = k_0 \exp(-\alpha N^\beta)$ where β should be 0.5 and α is a constant on the order of unity [15]. Similar arguments also lead to the same scaling for RNA folding as well [16]. The essence of the argument for expecting a sub-linear dependence of the barrier height on N can be understood by noting that the driving force for folding proteins is to bury the hydrophobic residues whereas charged or polar residues are better accommodated by extending the chain. This intrinsic conflict produces some fraction of interactions that are favorable for folding and others that favor extended structures. Similarly, in RNA favorable base-pairing interactions and the hydrophobic nature of the bases tend to collapse RNA whereas the charged phosphate residues prefer extended structures. Thus, the distribution of activation free energy, $\Delta G_{UF}^\ddagger/k_B T$, between the folded and unfolded states is a sum of favorable and unfavorable terms. We expect from central limit theorem that the distribution of $\Delta G_{UF}^\ddagger/k_B T$ should be roughly Gaussian with dispersion $\langle (\Delta G_{UF}^\ddagger)^2 \rangle \sim N$. Thus, $\Delta G_{UF}^\ddagger/k_B T \sim N^\beta$ with $\beta = 1/2$.

Remarkably, for both proteins (see Fig. 7 in ref. [17]) and RNA (Fig. 1) the predicted dependence of

the folding rate on N is extremely well-described by the theoretical prediction. From the fits of theory to experiments, we find that the inverse of the prefactor for RNA, $k_0^{-1} = \tau_0 \approx 0.87 \mu s$, is almost six orders of magnitude larger than the transition state theory estimate of $h/k_B T \approx 0.16 ps$. The value of τ_0 , which coincides with the typical base pairing time [18], is hence the speed limit for RNA folding. The predicted value for τ_0 is close to the speed limit established for protein folding as well [19]. The common speed limit suggests that the initial events triggering folding (in all likelihood favorable loop formation leading to hairpin formation in RNA and nucleation in protein folding) may be similar.

Final Remarks: The illustrations here do not even come close to capturing the excitement of working at the interface between biology and physics. The prospects of unearthing general principles governing living systems using physics concepts have never been greater. As a result the ever evolving biological world is a perfect playground for physicists. As the applications grow more quantitative, it is natural that physicists will come to grips with crucial differences between the living and non-living matter. The notions of adaptation and evolution, which play a crucial role in living matter at all length scales, have to be integrated into the theoretical description of biological processes. It is clear that that in grappling with these problems, physicists, who are most definitely up to the task, will perpetually feel like kids in a candy store!

Acknowledgements: I am grateful to an extraordinary group of talented scientists who taught me a great deal during their involvement in biophysics research. These researchers, many of whom are in academic institutions world wide, are most responsible for my continued excitement in biophysics research. I am grateful to Krastan Blagoev for inviting me to write this piece and for useful comments. Our works have been generously supported by the National Sci-

ence Foundation for nearly twenty five years most recently through CHE 13-61946.

-
- [1] D. W. Thompson, *On Growth and Form* (Cambridge University Press, 1917).
 - [2] E. Schrodinger, *What is life* (Cambridge University Press, 1948).
 - [3] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. USA **87**, 3526 (1990).
 - [4] C. B. Anfinsen, Science **181**, 223 (1973).
 - [5] I. Baskakov, G. Legname, S. Prusiner, and F. Cohen, J. Biol. Chem. **276**, 19687 (2001).
 - [6] S. V. Solomatin, M. Greenfeld, S. Chu, and D. Herschlag, Nature **463**, 681 (2010).
 - [7] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
 - [8] K. Dill, Biochemistry **24**, 1501 (1985).
 - [9] D. Thirumalai and S. A. Woodson, Acc. Chem. Res. **29**, 433 (1996).
 - [10] T. Kirkpatrick and D. Thirumalai, Rev. Mod. Phys. **86**, xxx (2014).
 - [11] C. J. Camacho and D. Thirumalai, Phys. Rev. Lett. **71**, 2505 (1993).
 - [12] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, Proc. Natl. Acad. Sci. **109**, 16155 (2012).
 - [13] M. S. Li, D. K. Klimov, and D. Thirumalai, Phys. Rev. Lett. **93**, 268107 (2004).
 - [14] C. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. **90**, 6369 (1993).
 - [15] D. Thirumalai, J. Phys. I (Fr.) **5**, 1457 (1995).
 - [16] C. Hyeon and D. Thirumalai, Biophys. J. **102**, L11 (2012).
 - [17] K. Dill, K. Ghosh, and J. Schmit, Proc. Natl. Acad. Sci. USA **108**, 17876 (2011).
 - [18] D. Pörschke, O. Uhlenbeck, and F. Martin, Biopolymers **12**, 1313 (1973).
 - [19] J. Kubelka, J. Hofrichter, and W. A. Eaton, Curr. Opin. Struct. Biol. **14**, 76 (2004).